MEASURING BROADBAND AMERICA

PRACTICUM REVIEW

Authors: Neha Rawal – Attila Veres – Robert Moreira – Suman Basu

KEY FINDINGS

- Weekend prime time is the new Peak Period

KEY FINDINGS

- Weekend prime time is the new Peak Period
- Consistent Speed – accounting for variance

KEY FINDINGS

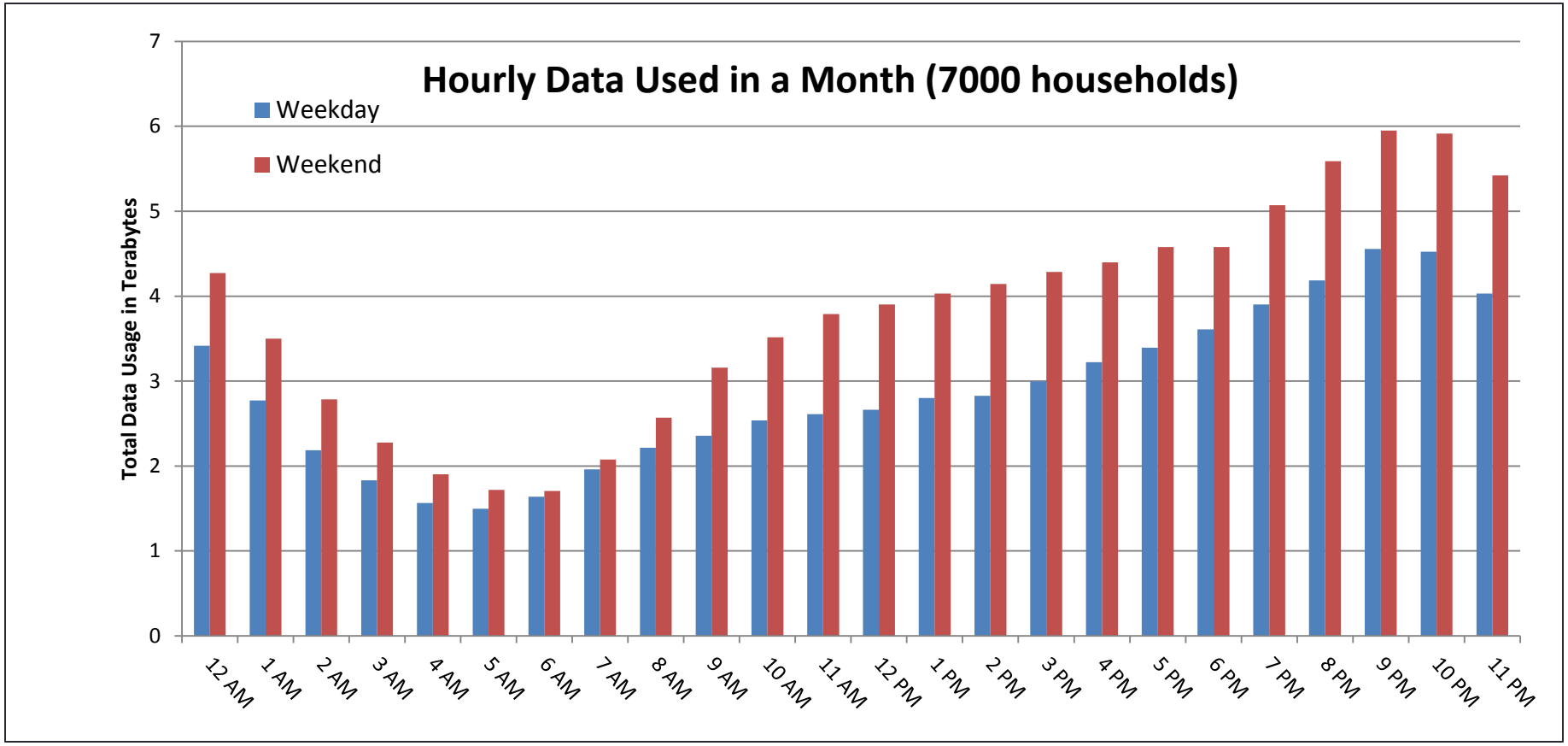
- Weekend prime time is the new Peak Period
- Consistent Speed – accounting for variance
- ISP alone does not tell the whole story

KEY FINDINGS

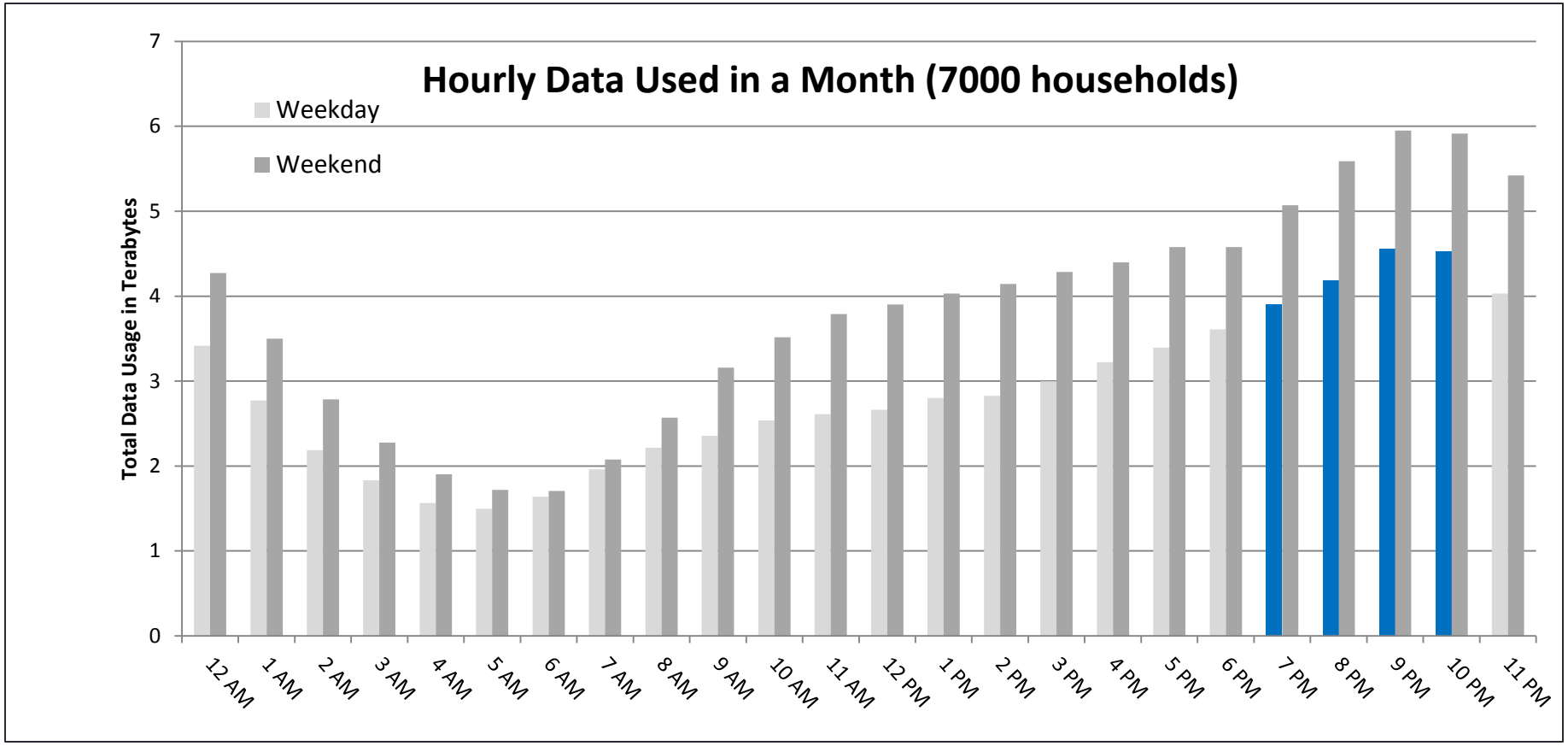
- Weekend prime time is the new Peak Period
- Consistent Speed – accounting for variance
- ISP alone does not tell the whole story
- If a picture says a thousand words – then an App says a million …

Data Usage

Data Usage Analysis

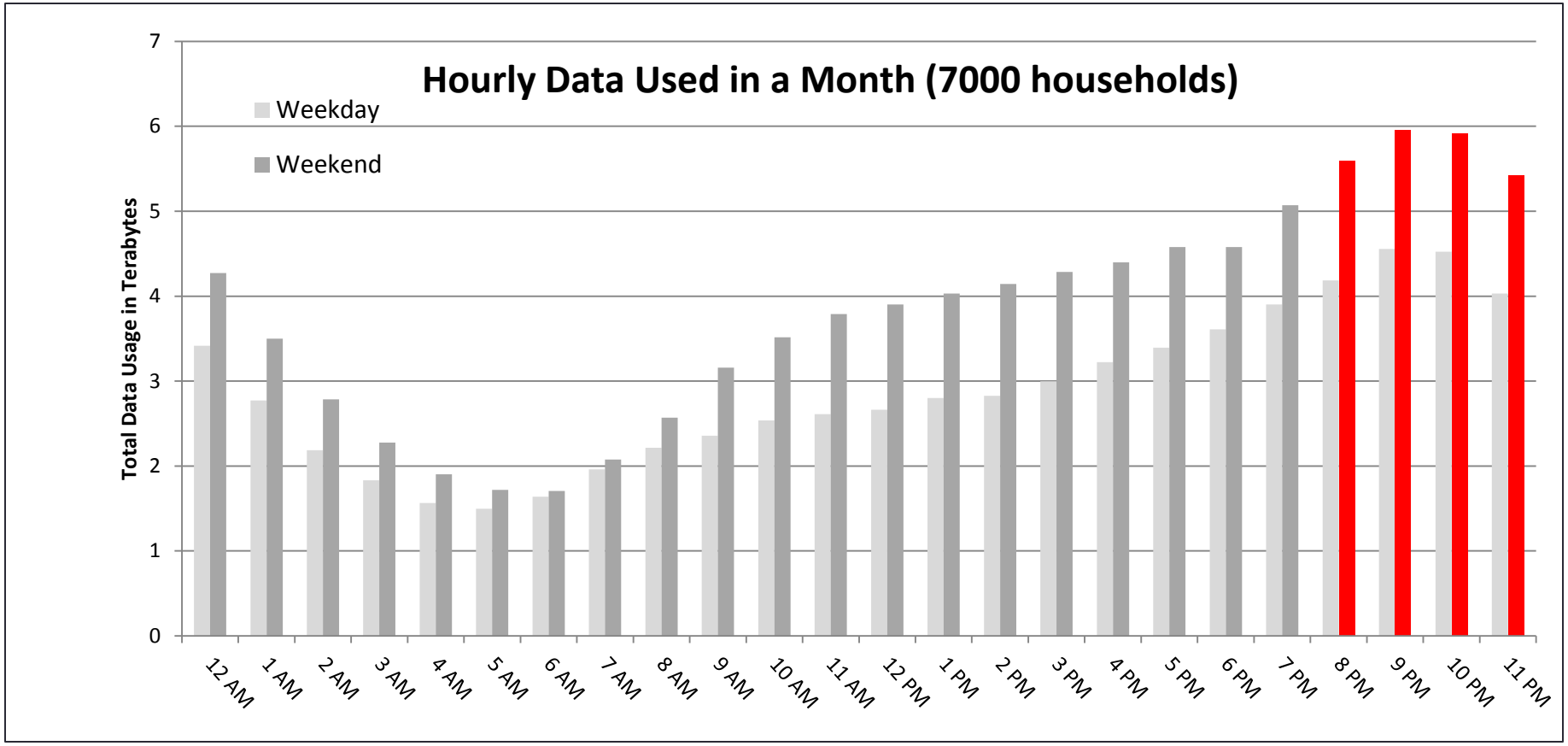


Data Usage Analysis



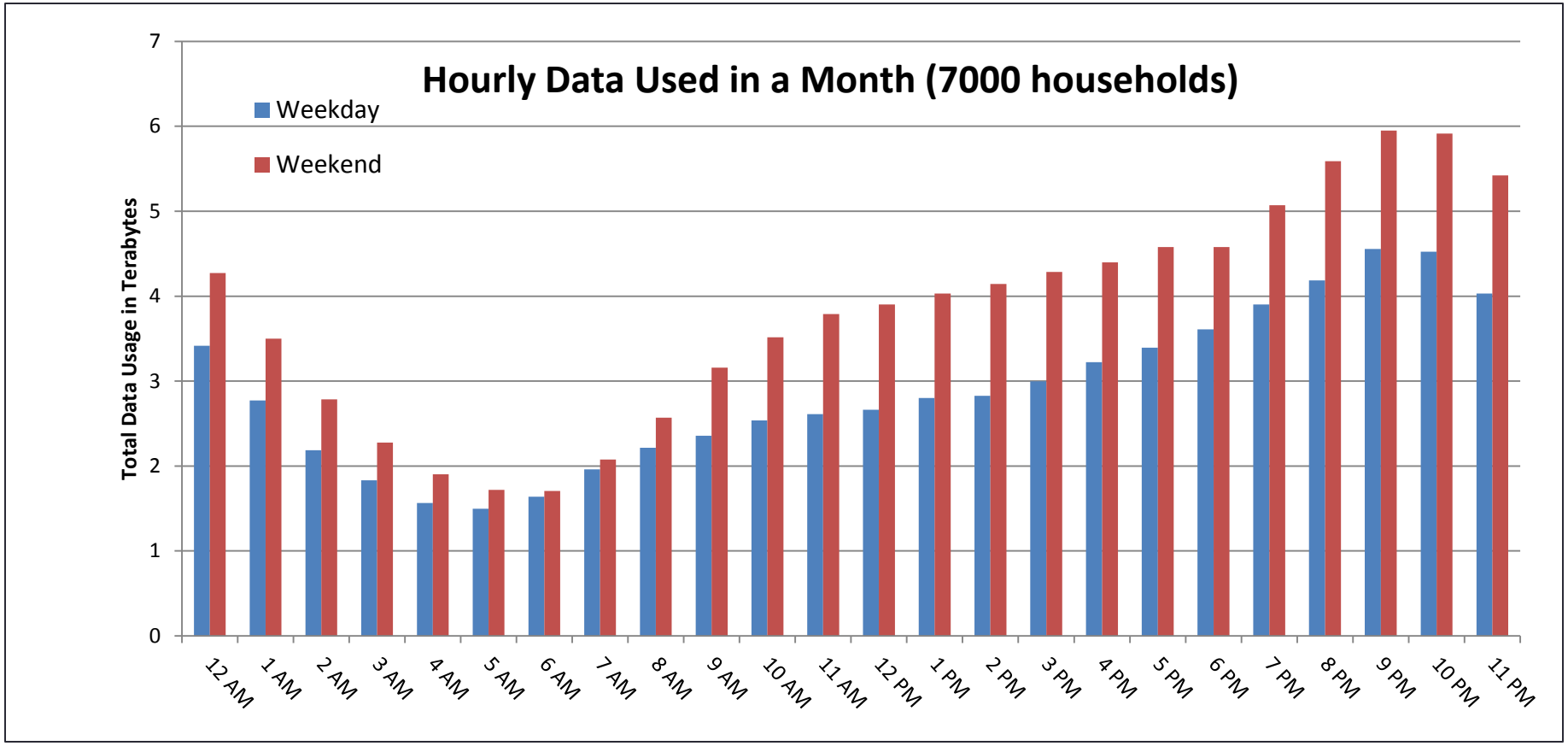
- Original Peak Period : Weekdays 7pm – 11pm

Data Usage Analysis



- Current Peak Period : Weekends 8pm – Midnight

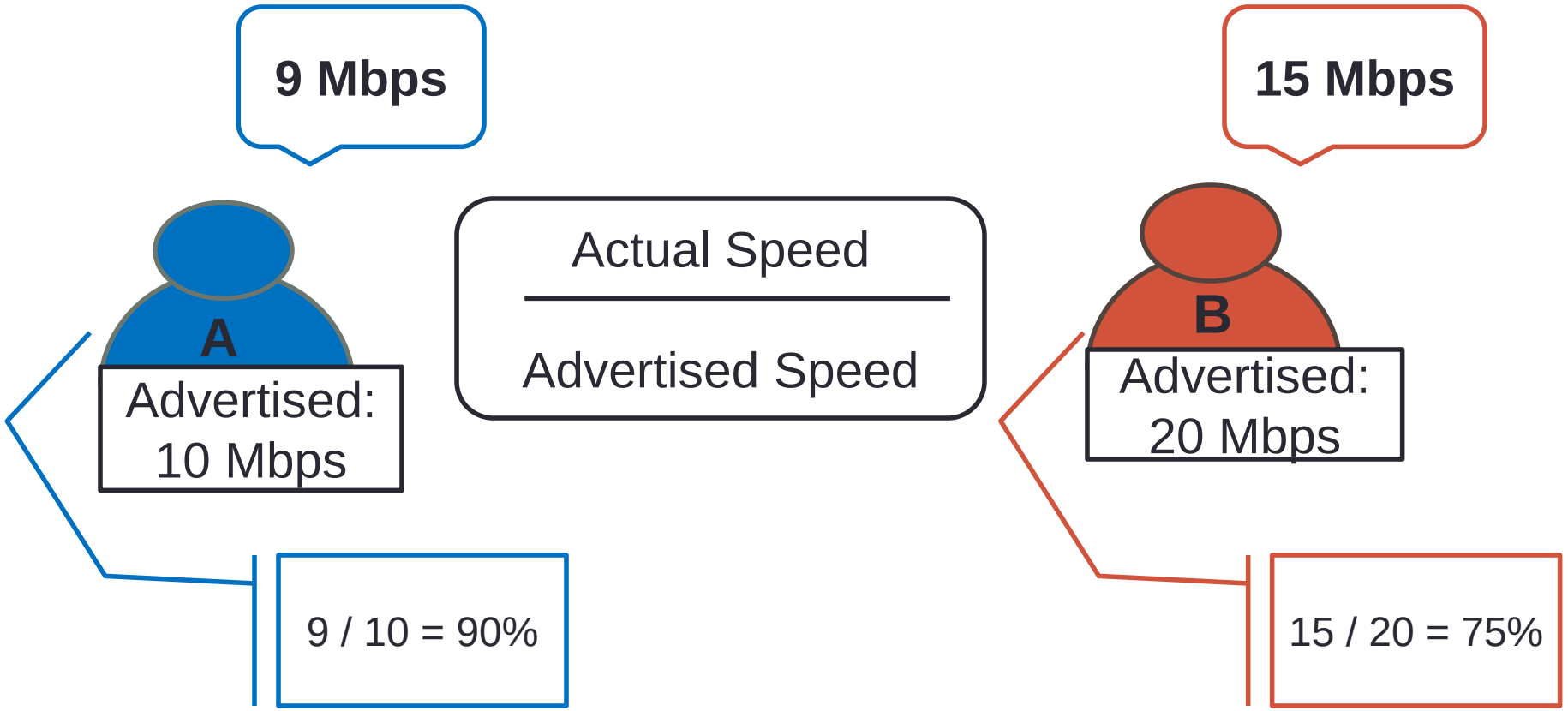
Data Usage Analysis



- Current Peak period : Weekends 8pm – Midnight
- 30% more data usage on average on weekends than workdays
- 50% more usage on average in evening hours than morning hours

Accounting for Variance

What's in a Performance Metric ?

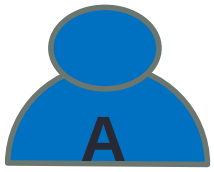


Variation : Temporal

Time

4 AM	8 AM	4 PM	8 PM
130%	100%	60%	50%

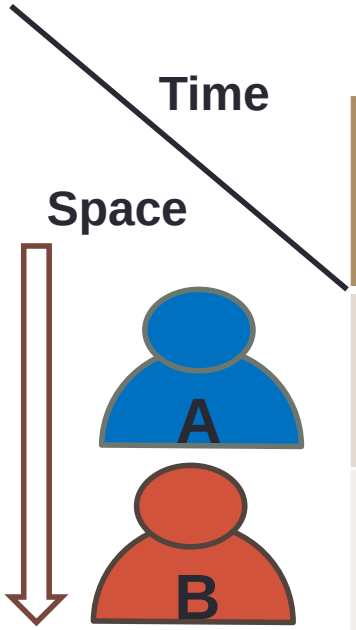
Actual / Advertised Speeds



Aggregation at unit ID level

Average	Percentile 25th
85%	60%

Variation : Spatial



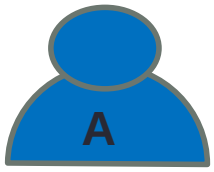
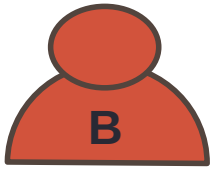
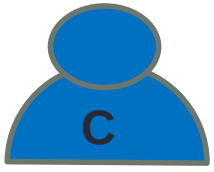
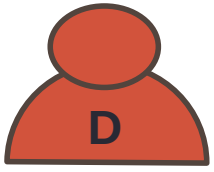
4 AM	8 AM	4 PM	8 PM
130%	100%	60%	50%
100%	90%	80%	70%

Aggregation at unit ID level

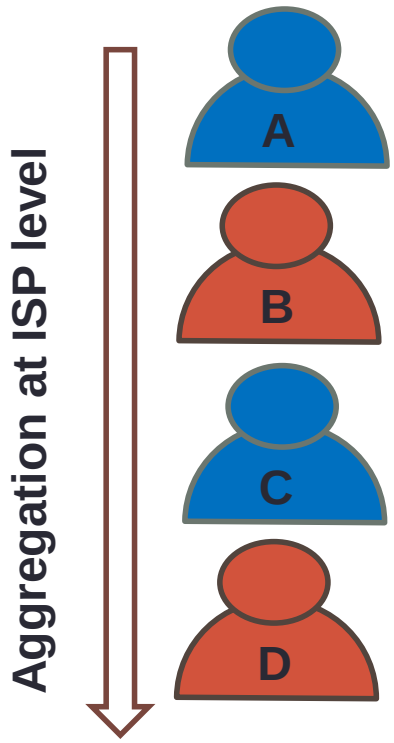
Average	Percentile 25th
85%	60%
85%	80%

Variation : Temporal

Aggregation at unit ID level

Time	4 AM	8 AM	4 PM	8 PM	Average	Percentile 25th
 A	130%	100%	60%	50%	85%	60%
 B	100%	90%	80%	70%	85%	80%
 C	110%	100%	100%	90%	100%	100%
 D	90%	70%	50%	30%	60%	50%

Variation : Spatial



Aggregation at unit ID level

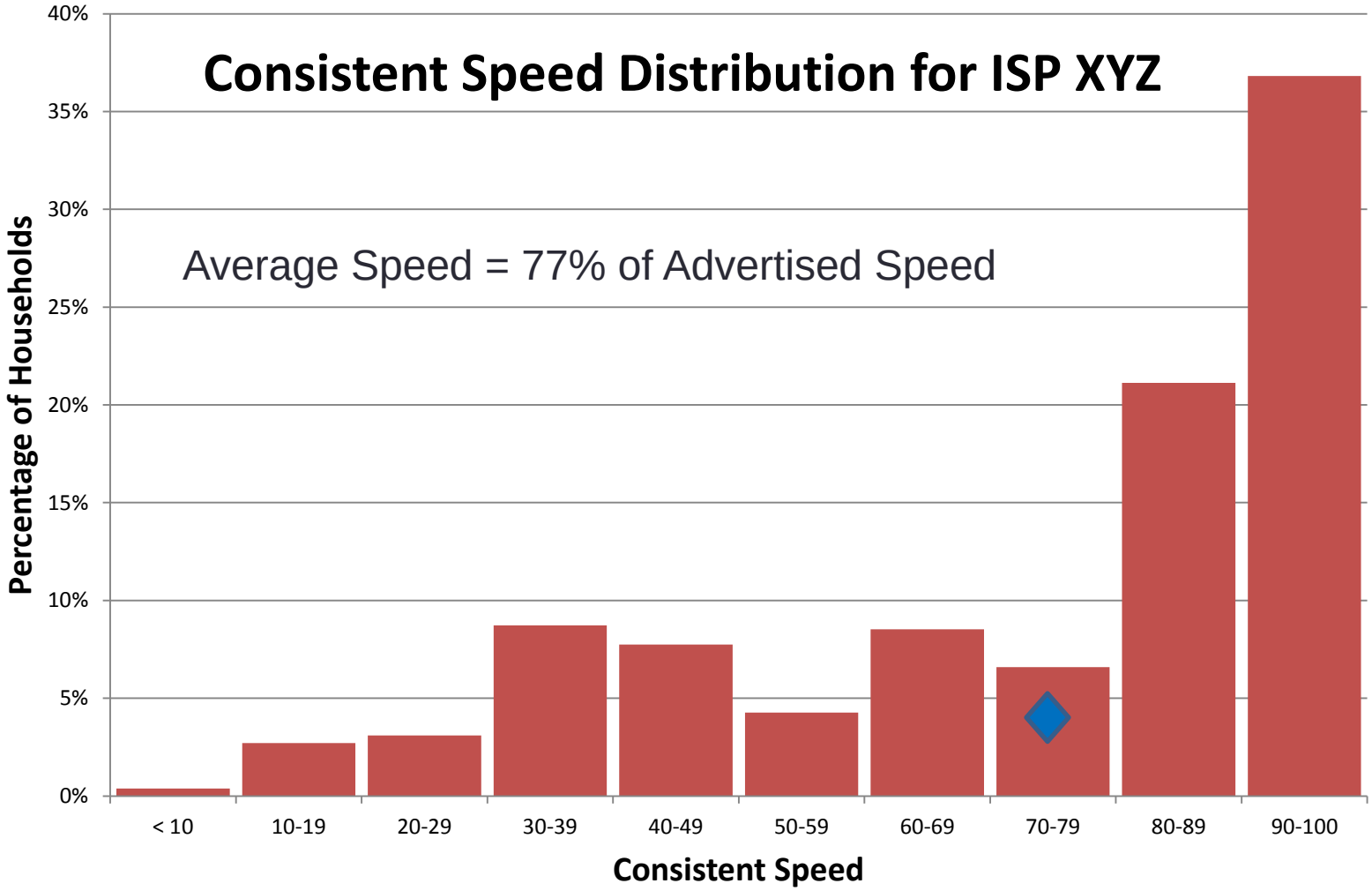
Average	Percentile 25th
85%	60%
85%	80%
100%	100%
60%	50%
82.5%	60%

ISP XYZ

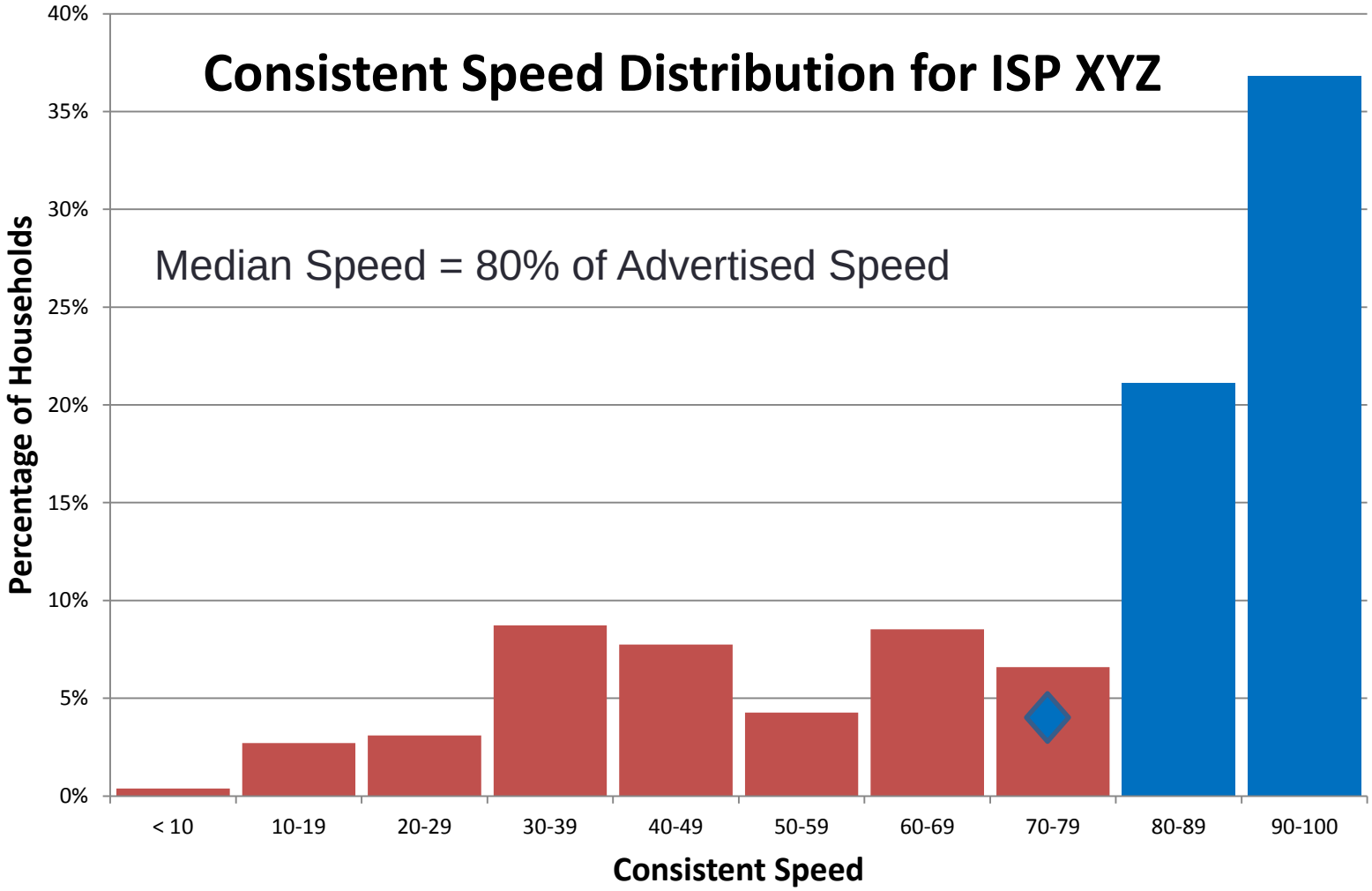
25th Percentile Consistent Speed (Speed 75 / 75)

At least 3 out of 4 households get 60% (or better) of advertised speed
3 out of every 4 hours

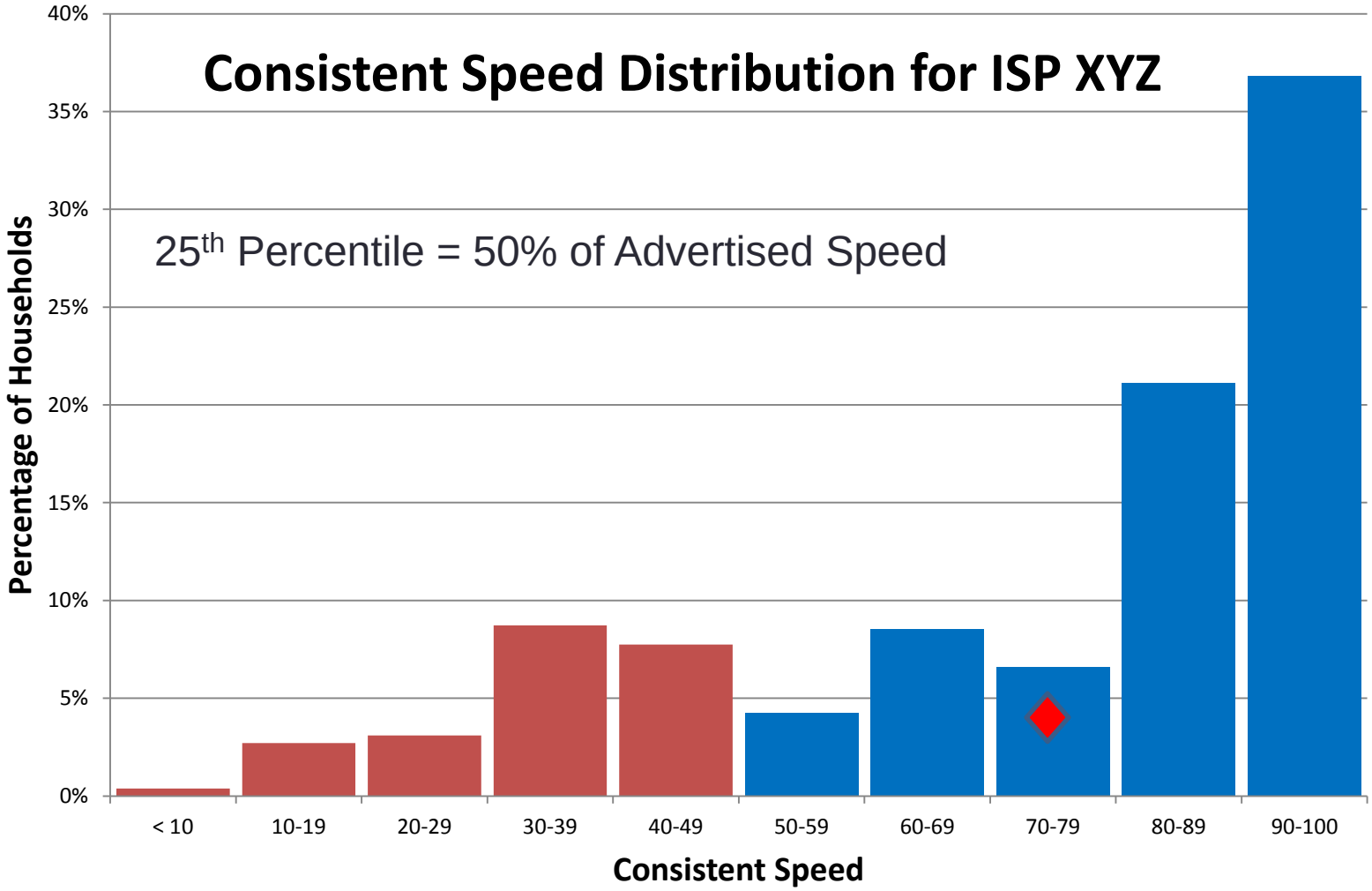
Averages vs. Percentiles



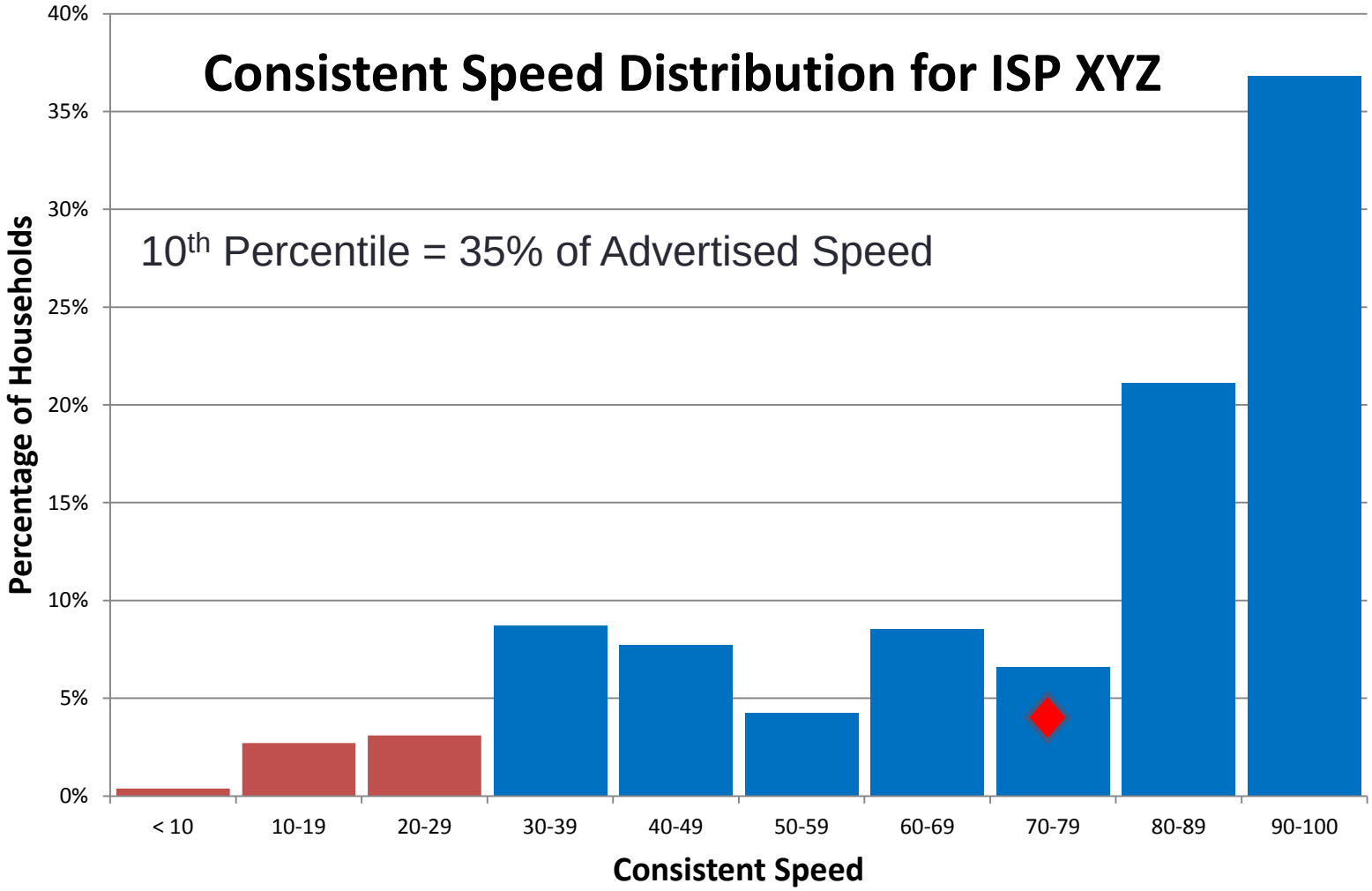
Averages vs. Percentiles



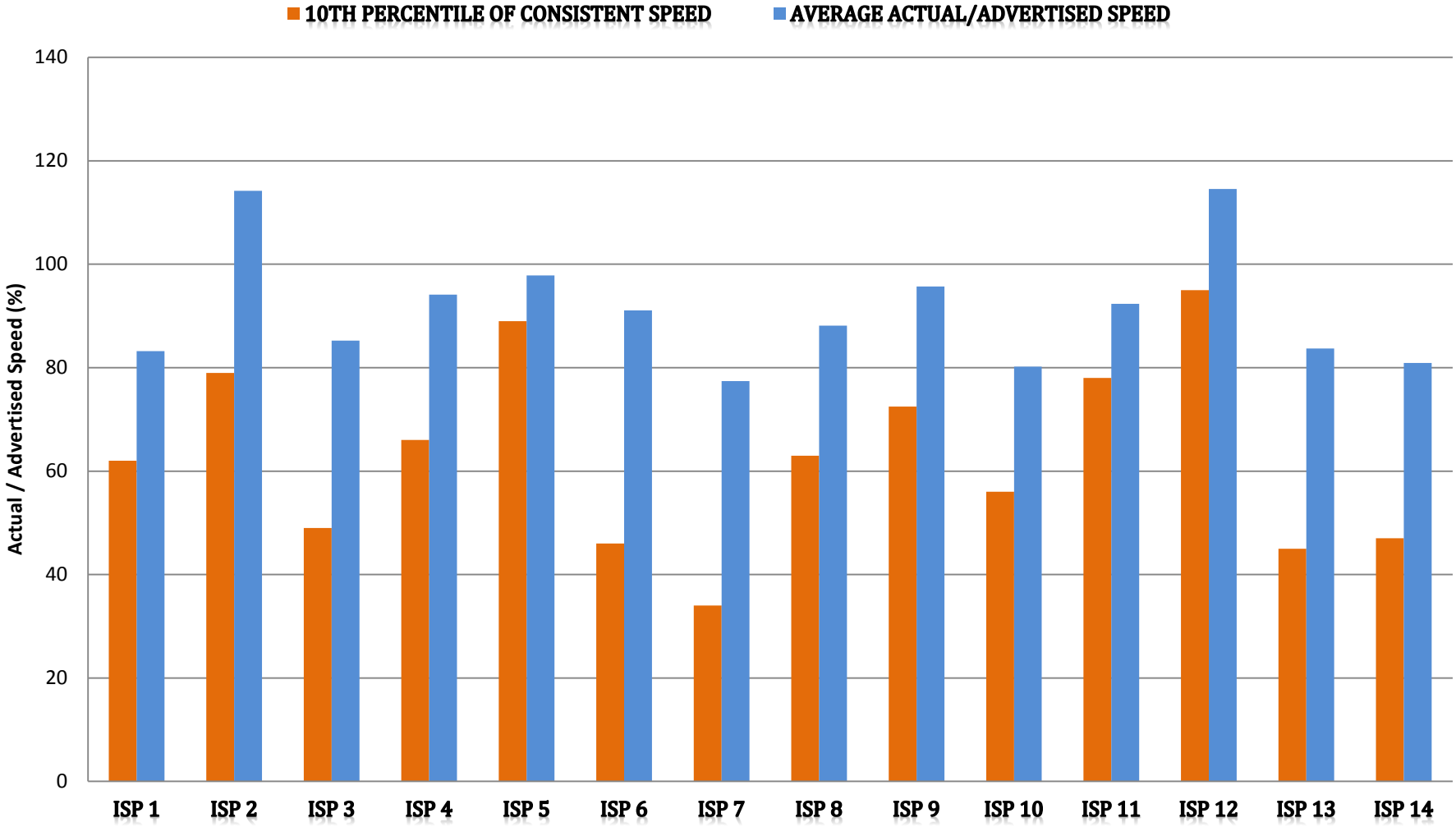
Averages vs. Percentiles



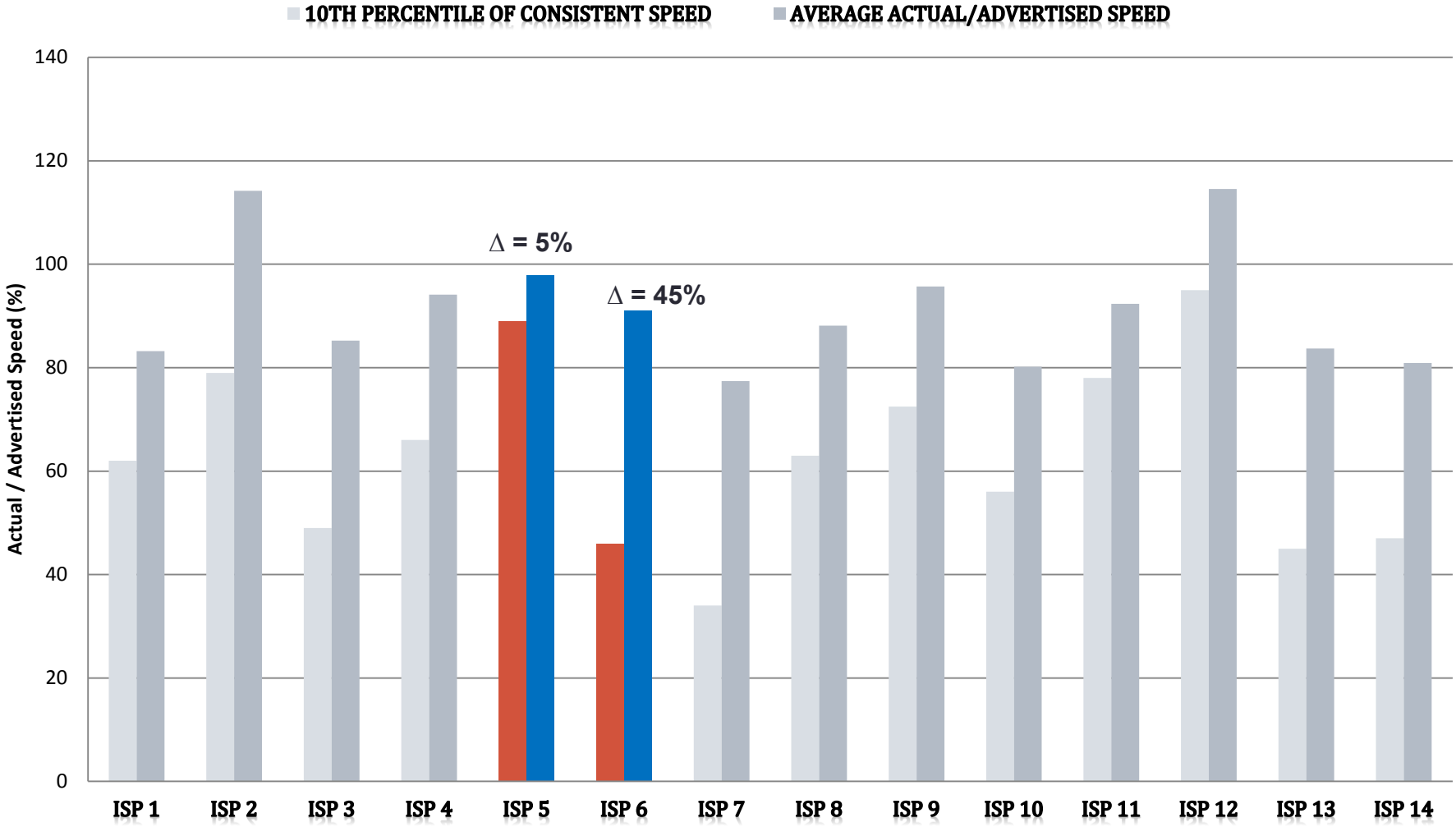
Averages vs. Percentiles



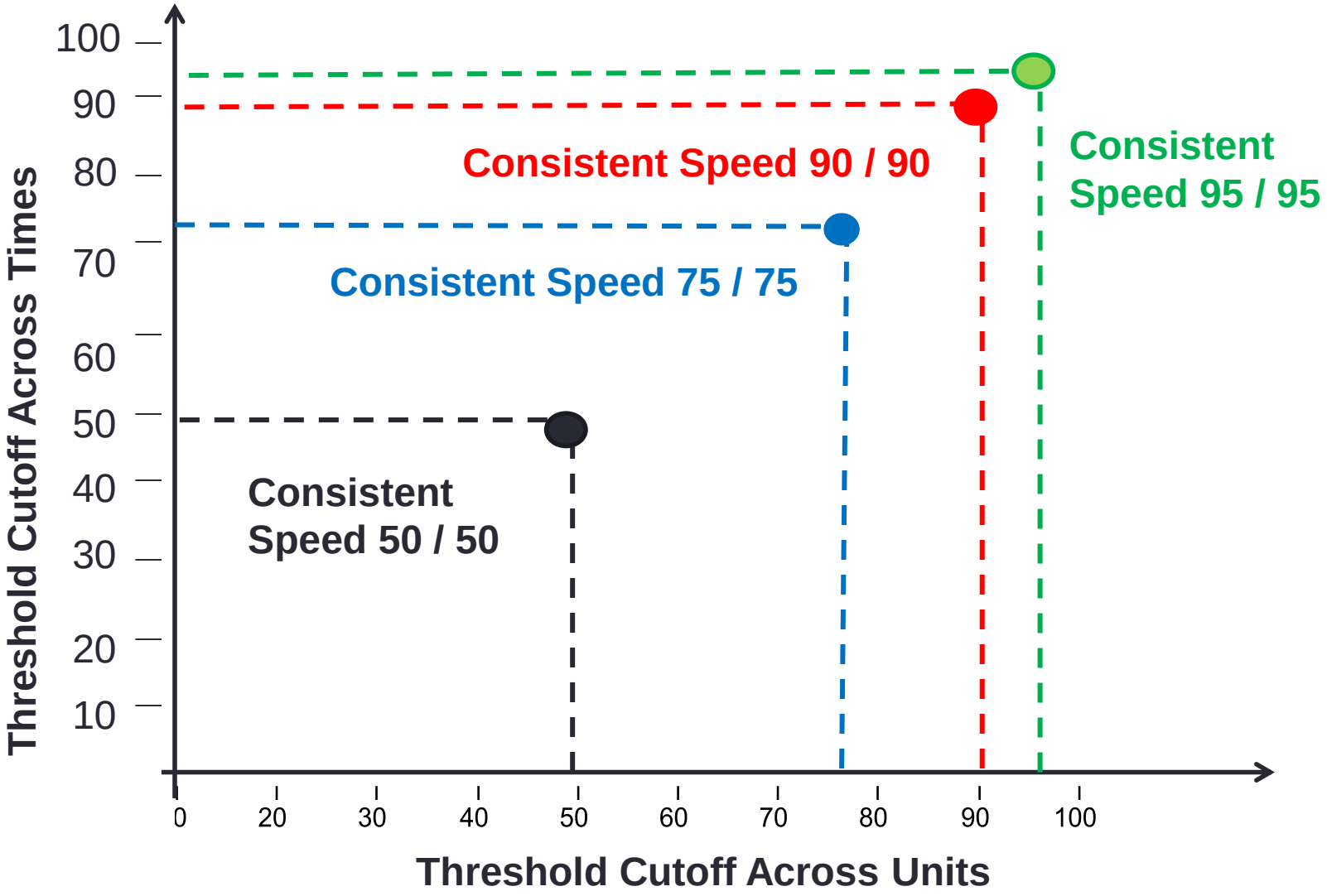
CONSISTENT SPEED VS. AVERAGE SPEED

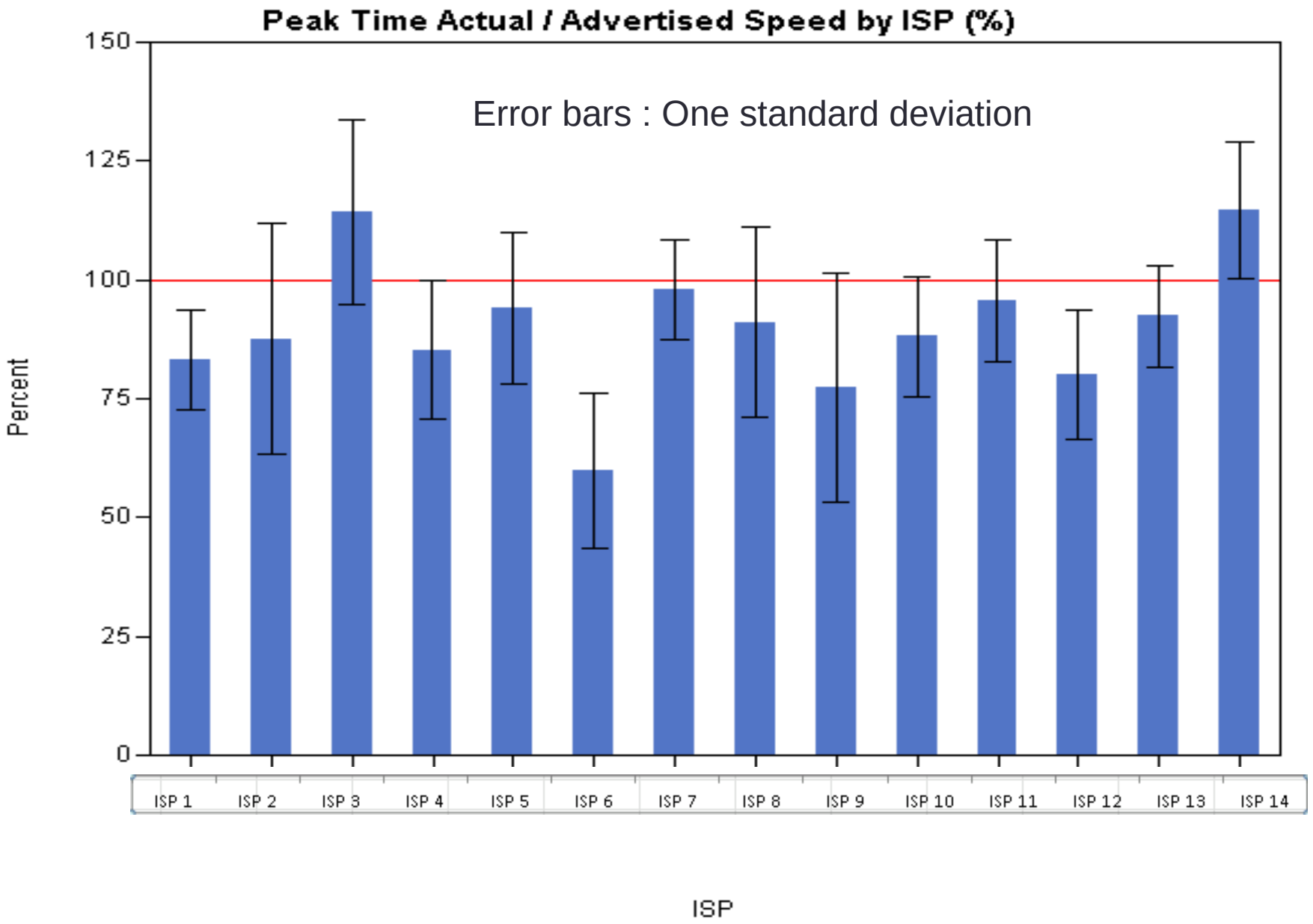


CONSISTENT SPEED VS. AVERAGE SPEED

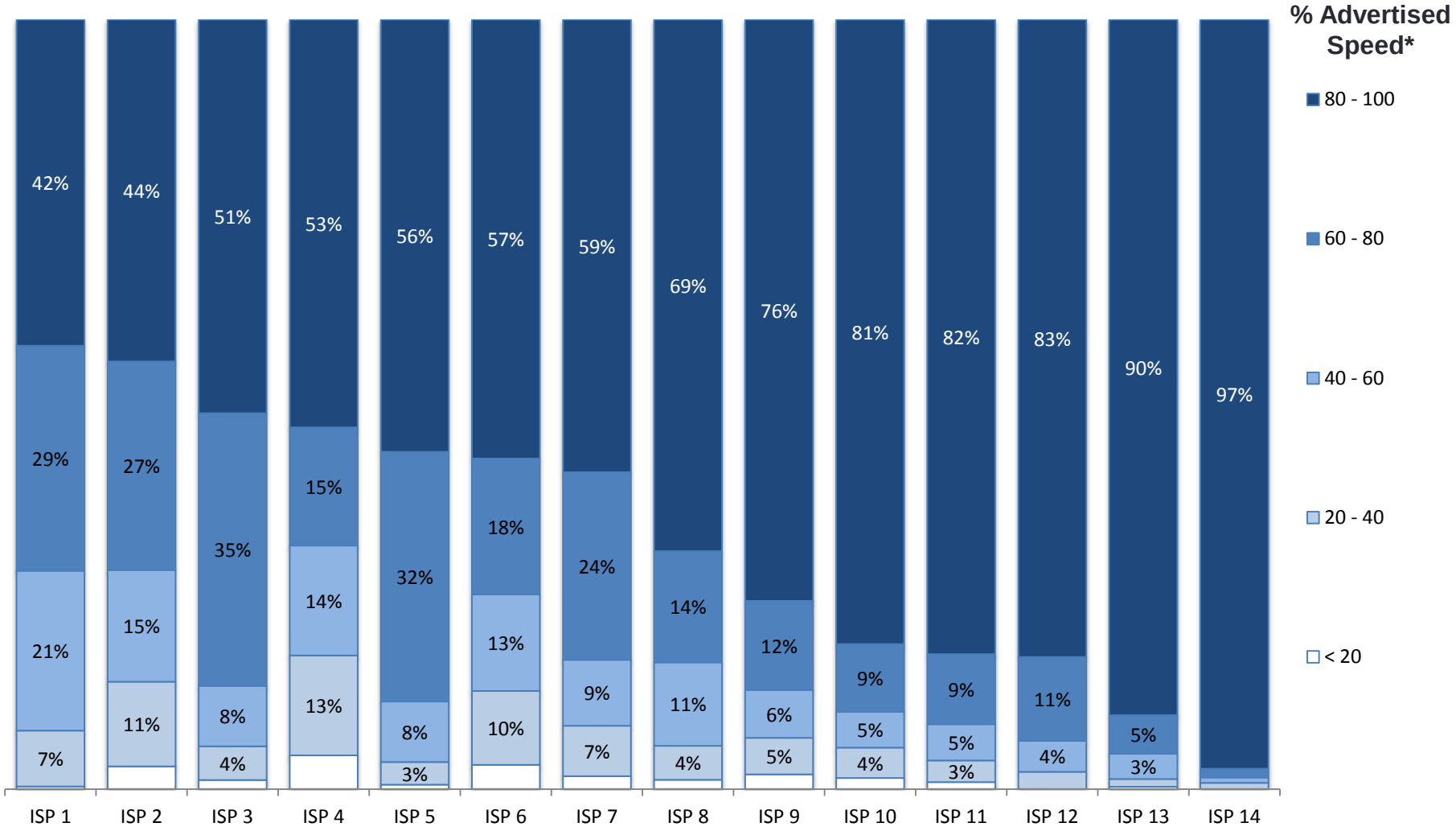


How the FCC Can Use Consistent Speed …




Peak Time Actual / Advertised Speed by ISP (%)
Error bars : One standard deviation
Percent
150
125
100
75
50
25
0
ISP 1
ISP 2
ISP 3
ISP 4
ISP 5
ISP 6
ISP 7
ISP 8
ISP 9
ISP 10
ISP 11
ISP 12
ISP 13
ISP 14
ISP

Distribution of Actual / Advertised Consistent Speed across ISPs



* 90% of the times

Variable Importance

Variable Importance

- Goal: What should the FCC focus on measuring in order to improve broadband in America?

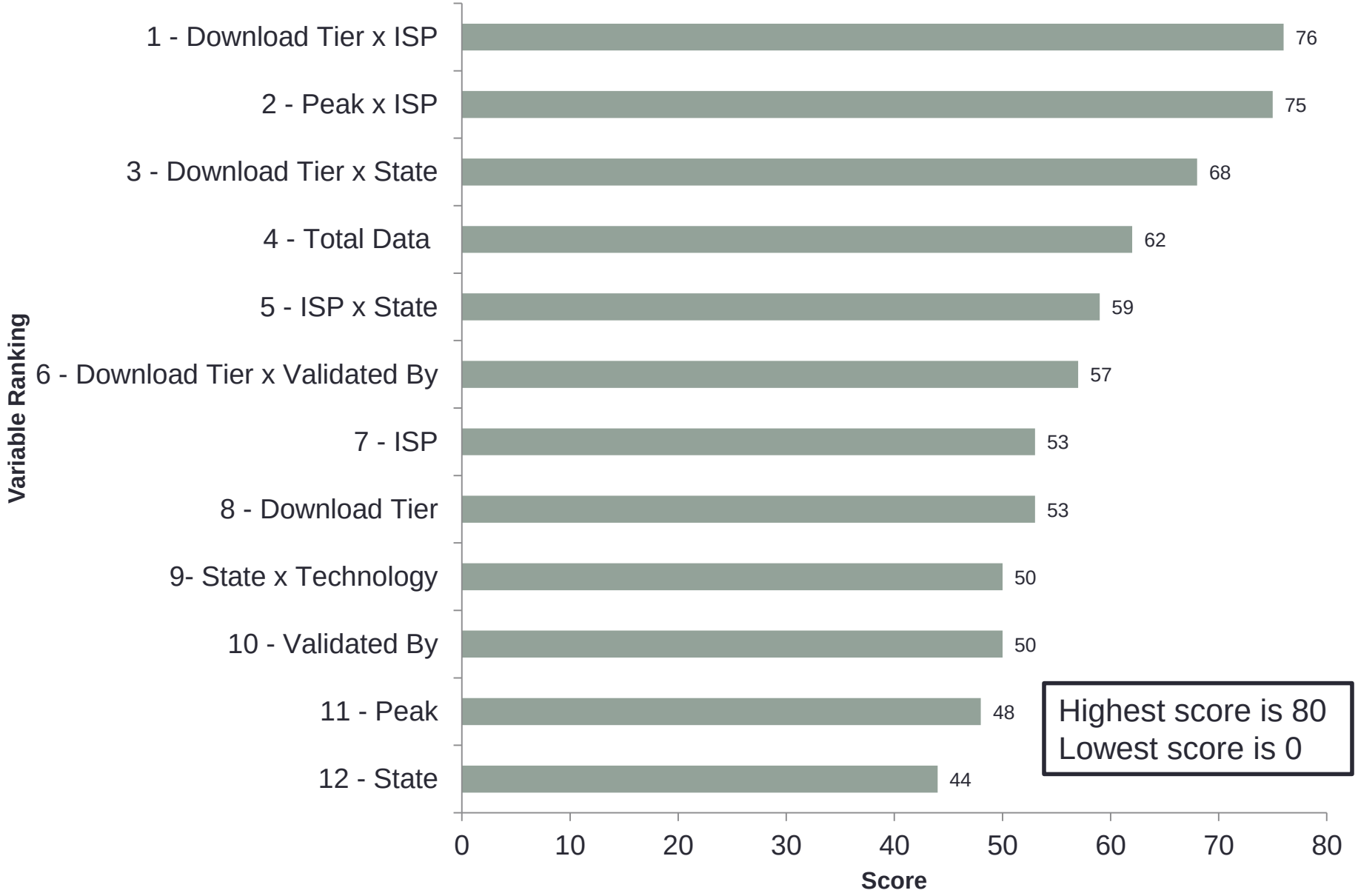
Variable Importance

- Goal: What should the FCC focus on measuring in order to improve broadband in America?

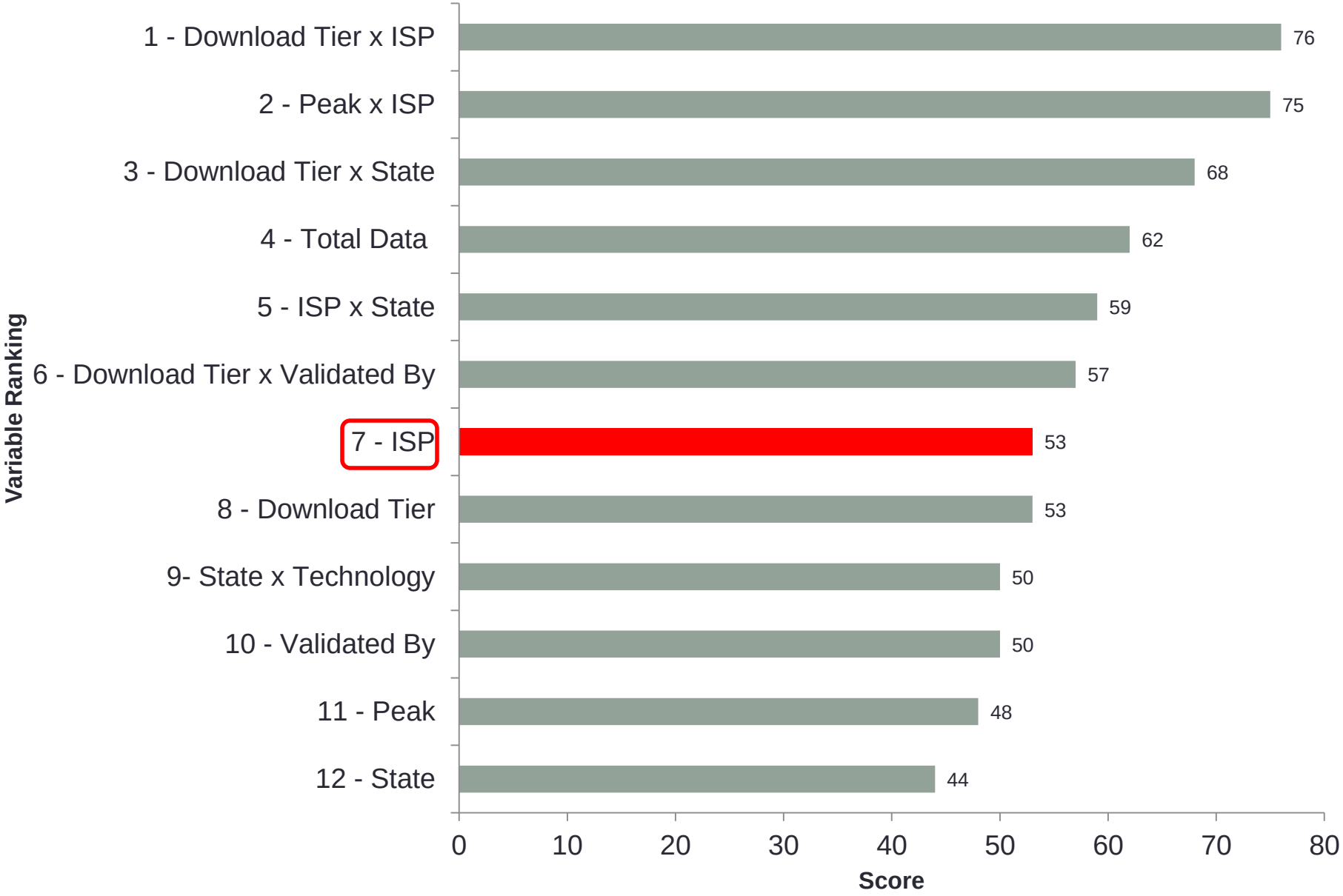
- What it is not: Model(s) to determine or infer parameter estimates – thousands of class variable levels made for difficult interpretation.

Explanatory Variable Scores

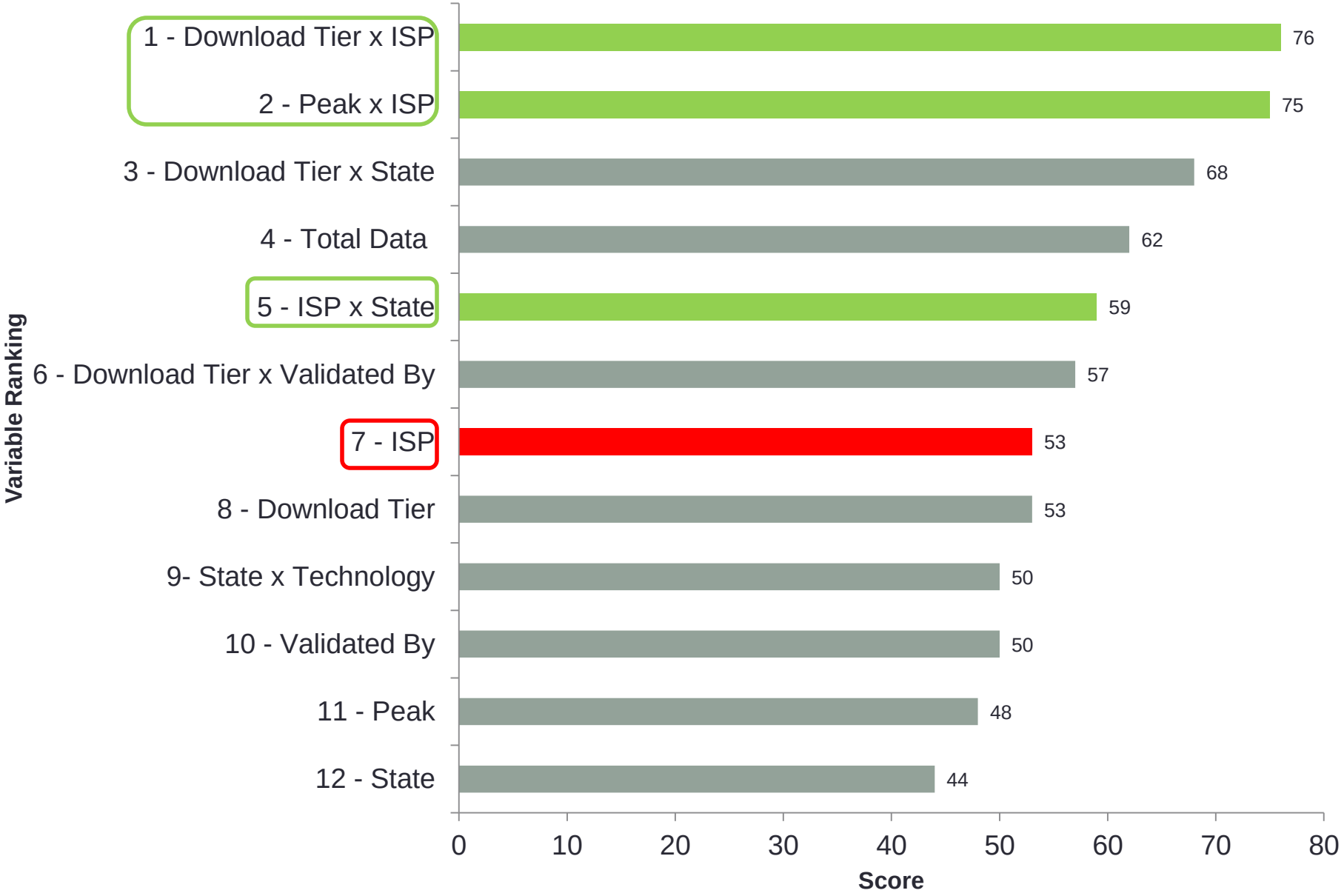
Explanatory Variable Score



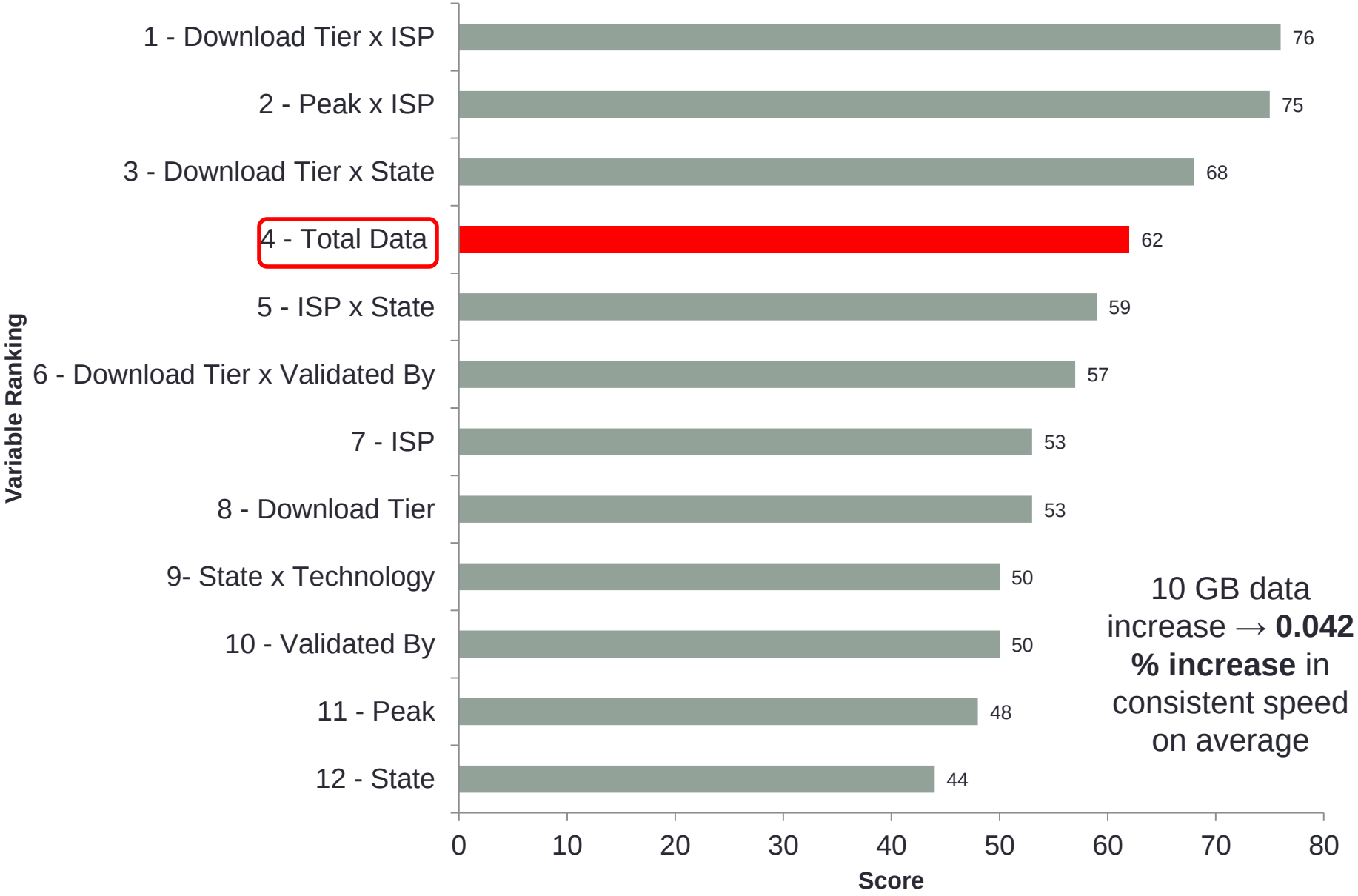
Explanatory Variable Score



Explanatory Variable Score

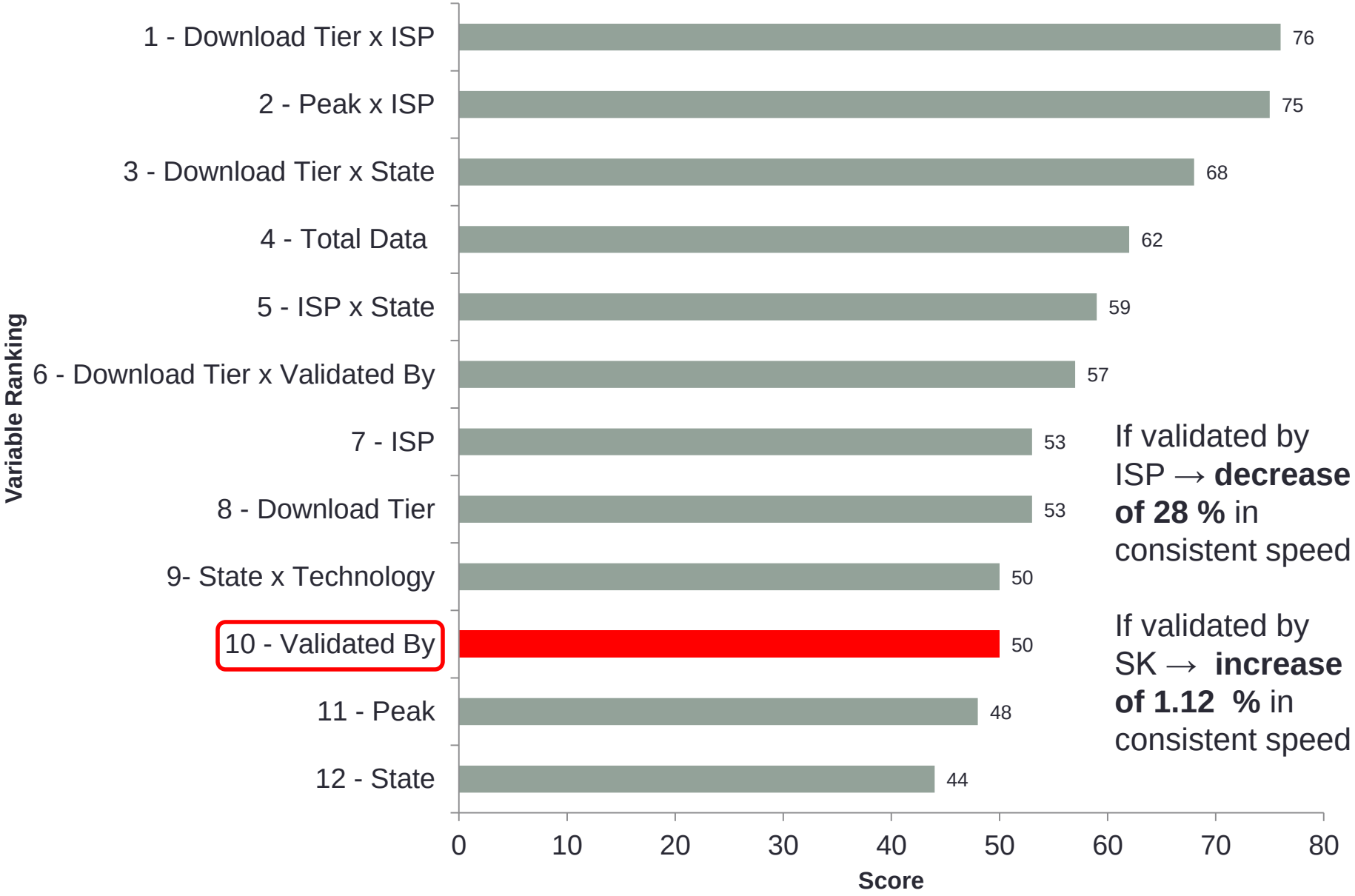


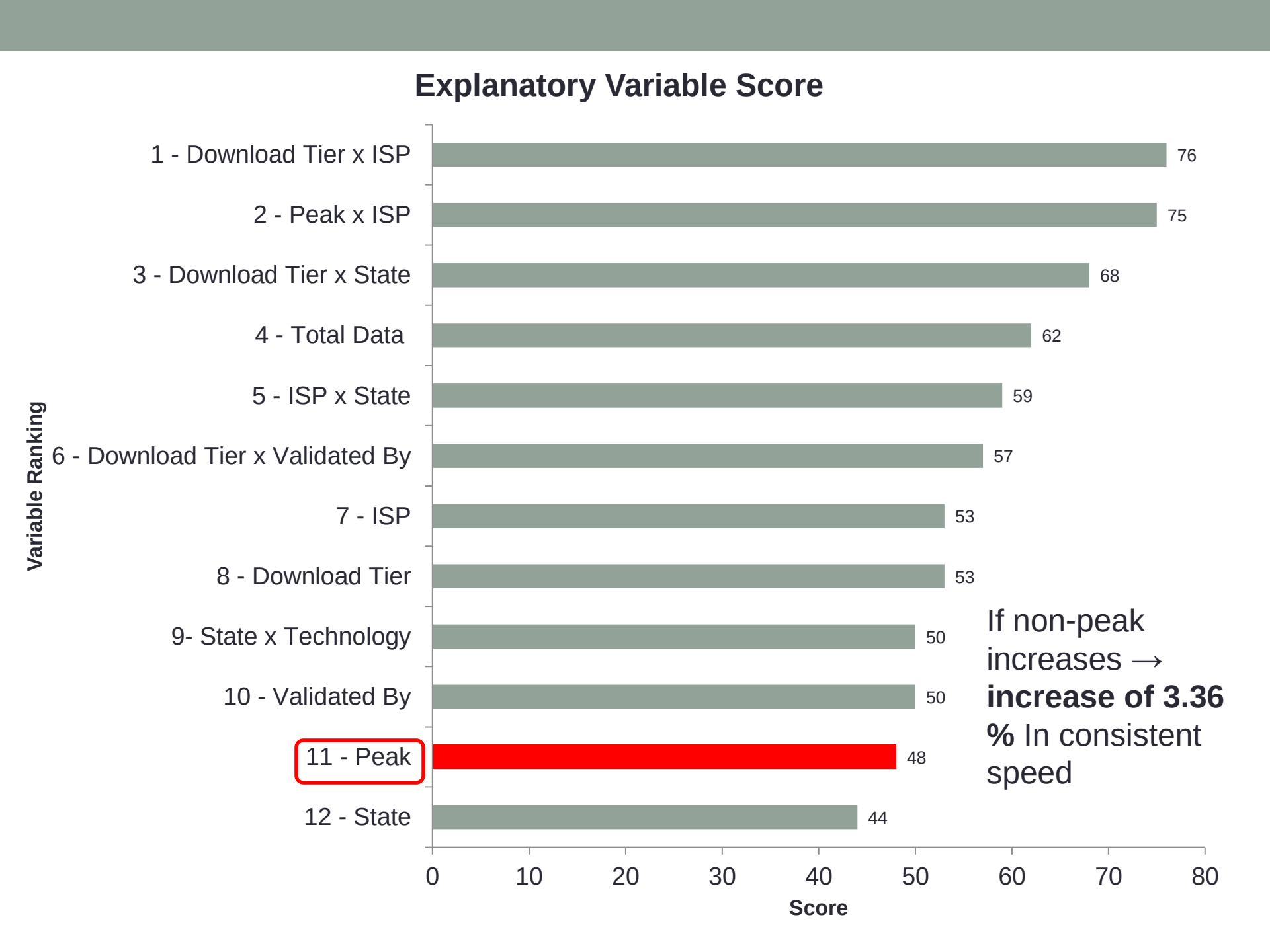
Explanatory Variable Score



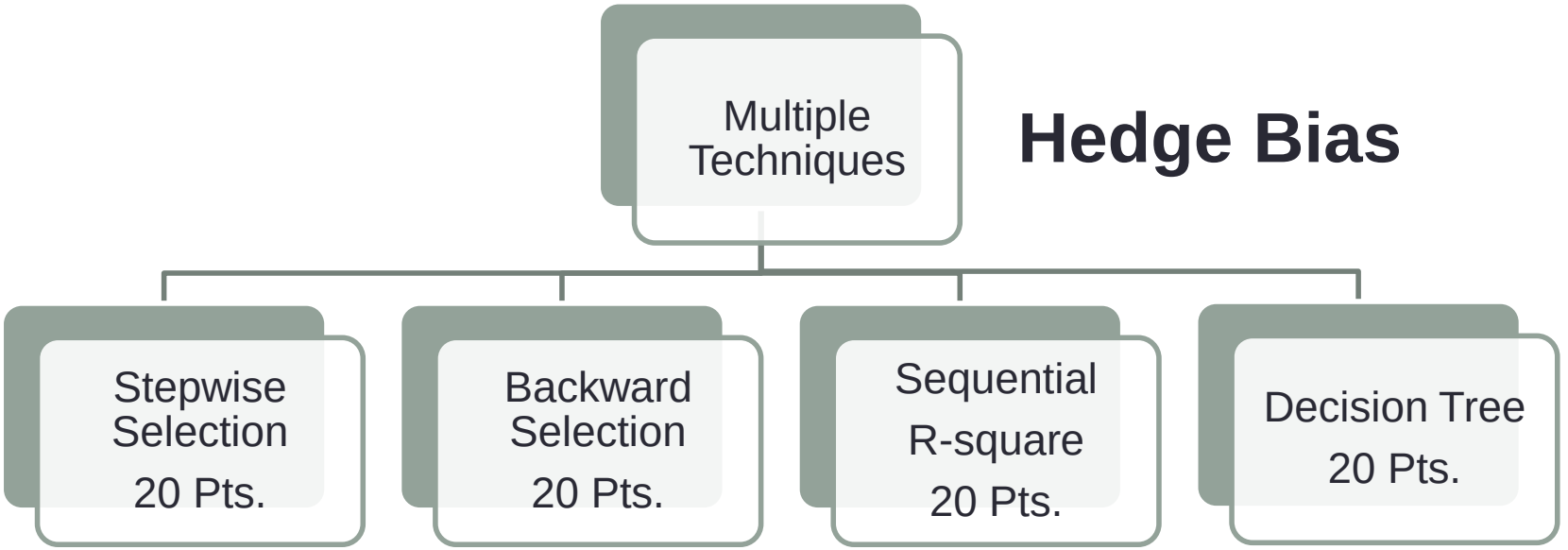
10 GB data increase → **0.042 % increase** in consistent speed on average

Explanatory Variable Score



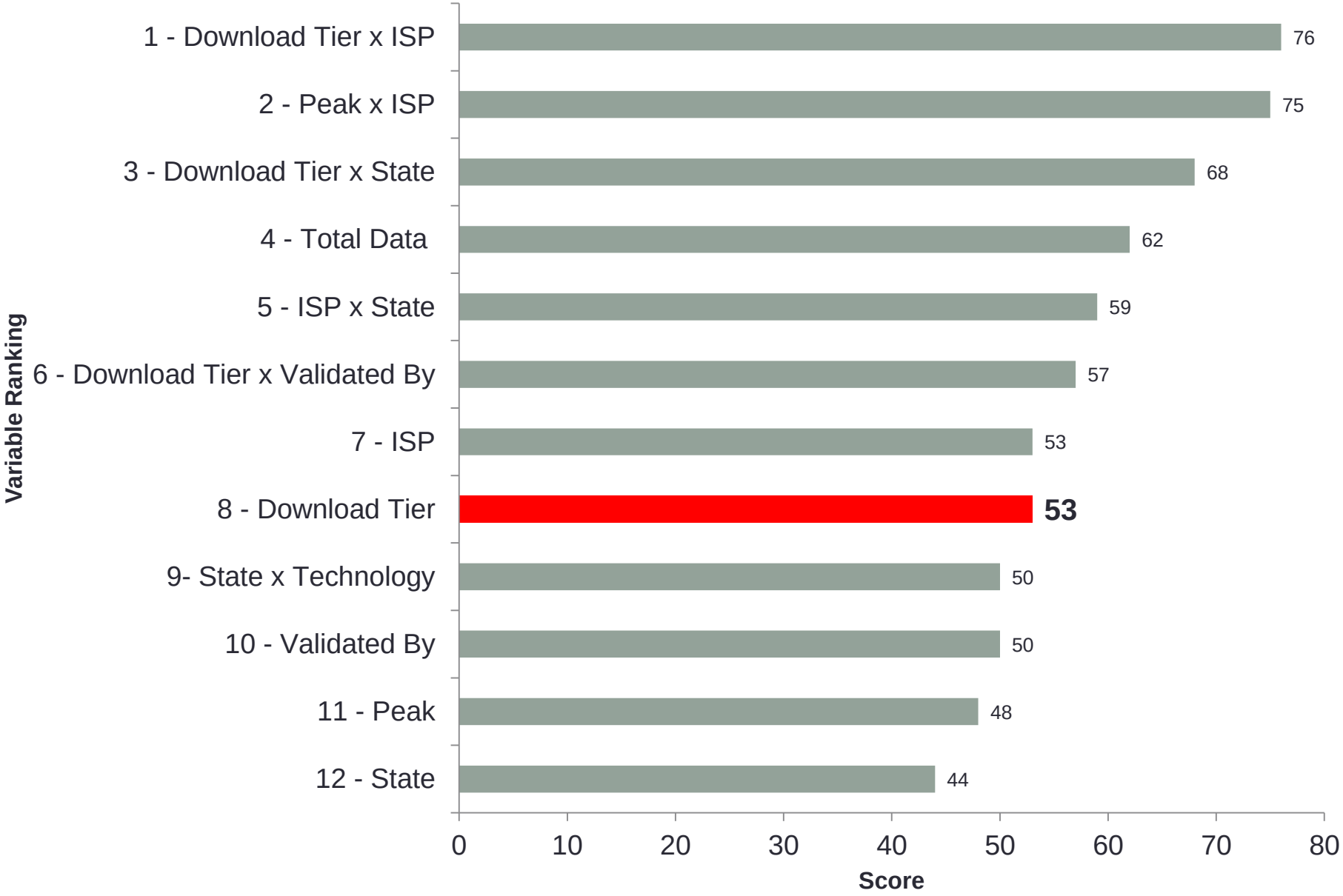

Explanatory Variable Score
Variable Ranking
1 - Download Tier x ISP
2 - Peak x ISP
3 - Download Tier x State
4 - Total Data
5 - ISP x State
6 - Download Tier x Validated By
7 - ISP
8 - Download Tier
9- State x Technology
10 - Validated By
11 - Peak
12 - State
76
75
68
62
59
57
53
53
50
50
48
44
0
10
20
30
40
50
60
70
80
Score
If non-peak increases → **increase of 3.36 %** In consistent speed

Variable Importance: Methodology



CUSTOM SCORING SYSTEM

Explanatory Variable Score



Variable Importance: Methodology

- Variable Selection Rank: 7 → 14 points
- Decision Tree Rank: not selected → 0 points
- Stepwise Regression Rank: 2 → 19 points
- Backward Regression: 20 points

- Score: 14+0+19+20 = 53

Variable Importance: Conclusion

- Still room for improvement

Variable Importance: Conclusion

- Still room for improvement
- Total Data, new Peak, & interaction variables increase explanatory power

Variable Importance: Conclusion

- Still room for improvement
- Total Data, new Peak, & interaction variables increase explanatory power
- Thousands of levels make parameter inference difficult

Variable Importance: Conclusion

- Still room for improvement
- Total Data, new Peak, & interaction variables increase explanatory power
- Thousands of levels make parameter inference difficult
- Focusing on ISP alone does not paint the whole picture

Why an App ?

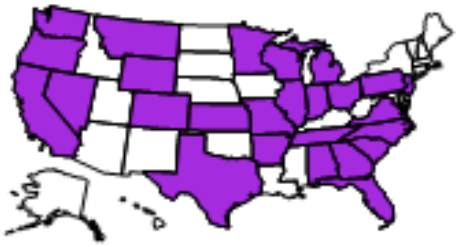
Presence of ISP by State

ISP 1

ISP 2

ISP 3

ISP 4



ISP 5

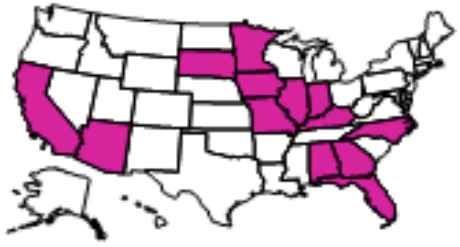
ISP 6

ISP 7

ISP 8

ISP 9

ISP 10



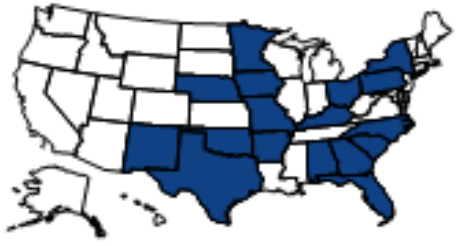
ISP 11

ISP 12

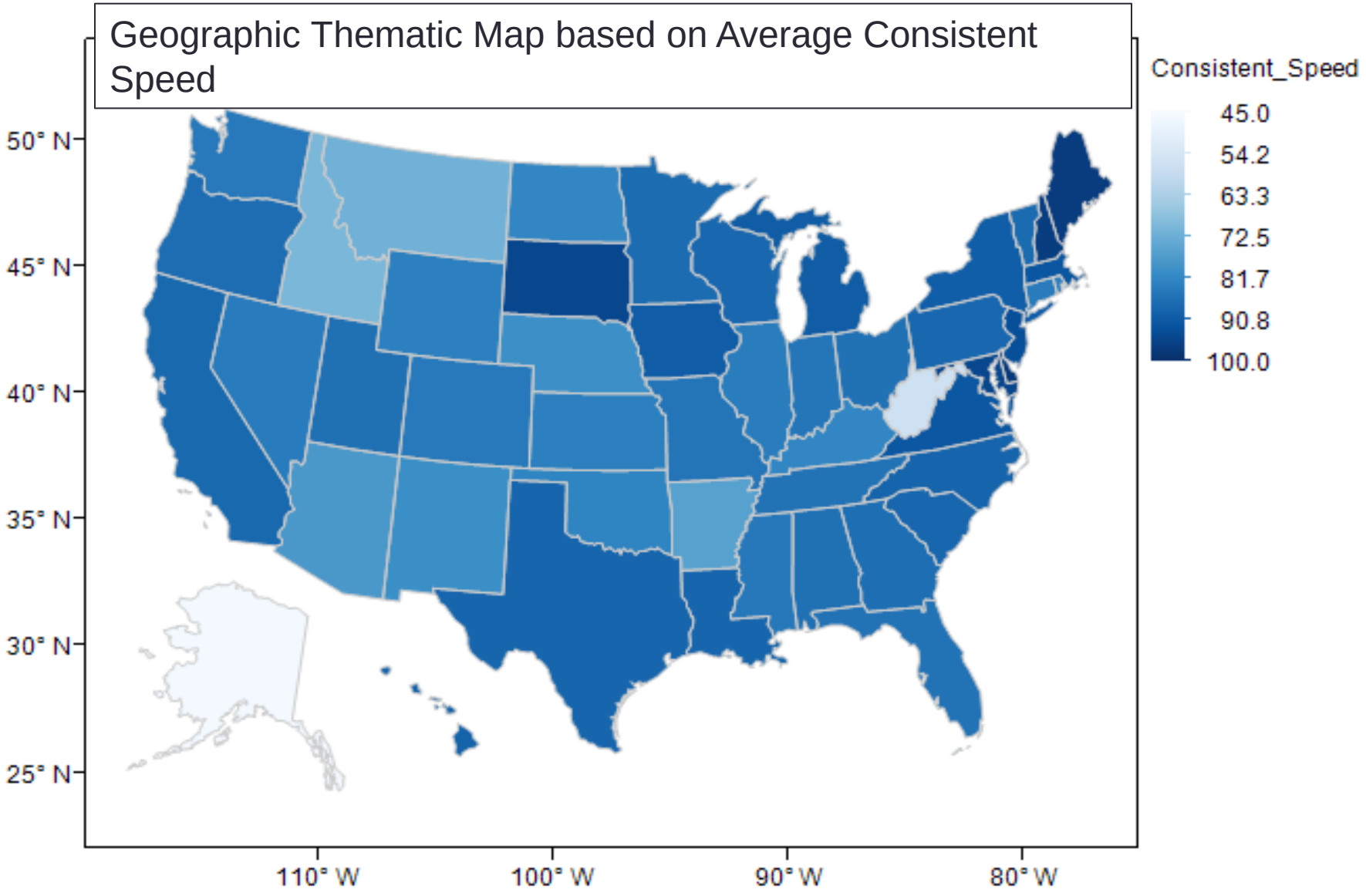
ISP 13

ISP 14

ISP 15



Based on sample data


Geographic Thematic Map based on Average Consistent Speed
Consistent_Speed
45.0
54.2
63.3
72.5
81.7
90.8
100.0
50° N
45° N
40° N
35° N
30° N
25° N
110° W
100° W
90° W
80° W

FCC Desktop Application

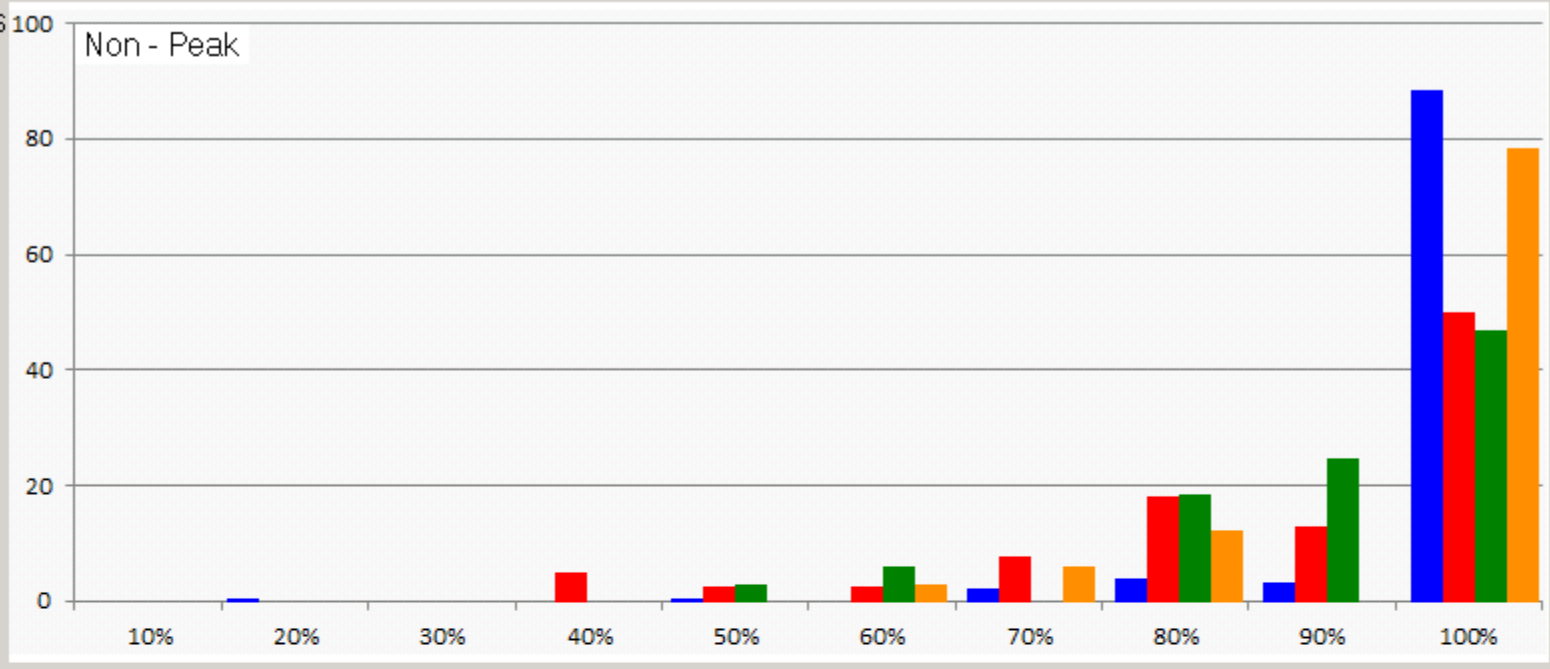
% Users

Select State

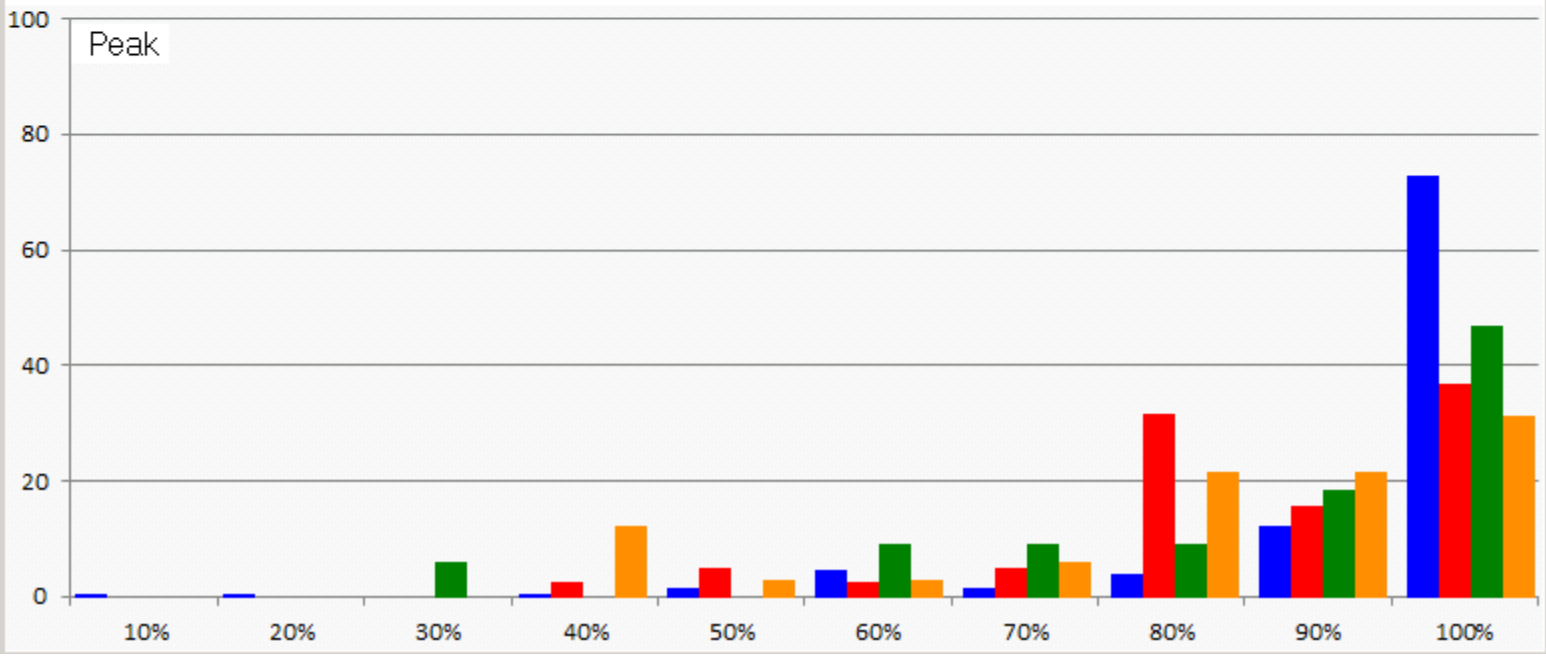
NY
NC
ND
OH
OK
OR
PA
SC
SD
TN
TX

Confirm

ISP 1
ISP 2
ISP 3
ISP 4



Consistent Speed >>



Preparing the Data

Used SAS to create and sort dataset and export it as an Access file

Access file is downloaded by users

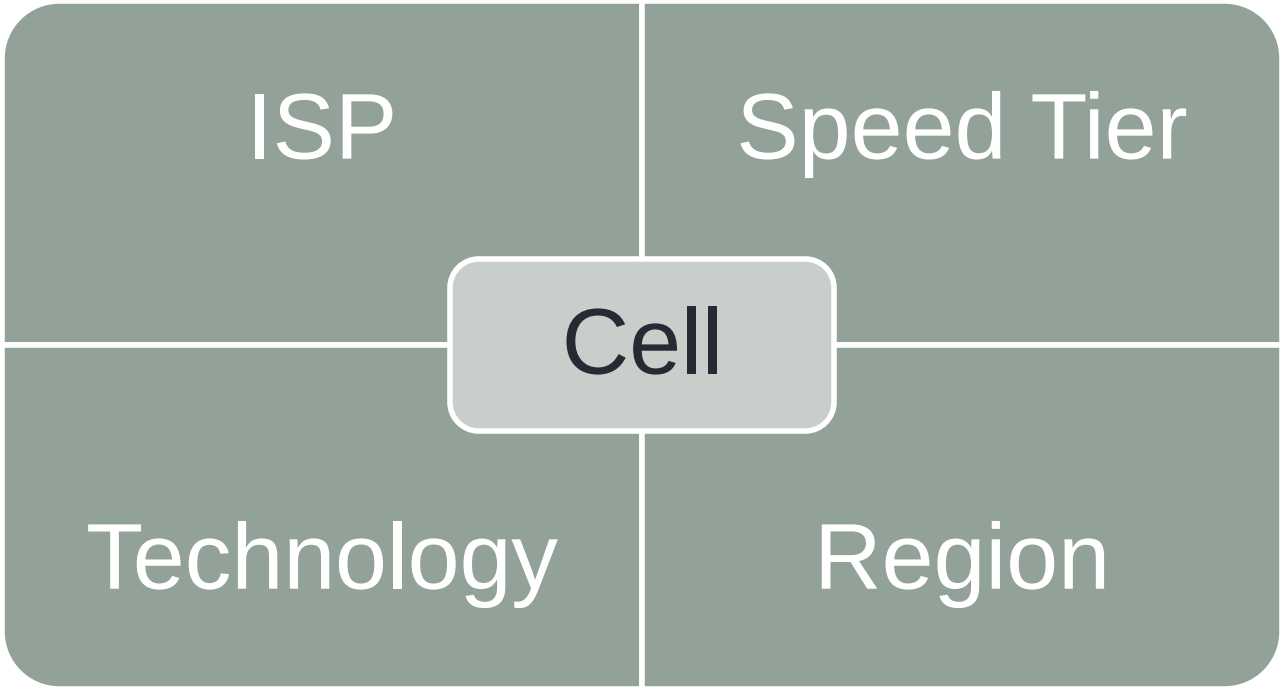
User links the Access file to the application on their PC

FCC creates the Access file and posts it online

The user downloads the Access file, and the desktop application and links them together

Validation & Review

Sampling Methodology

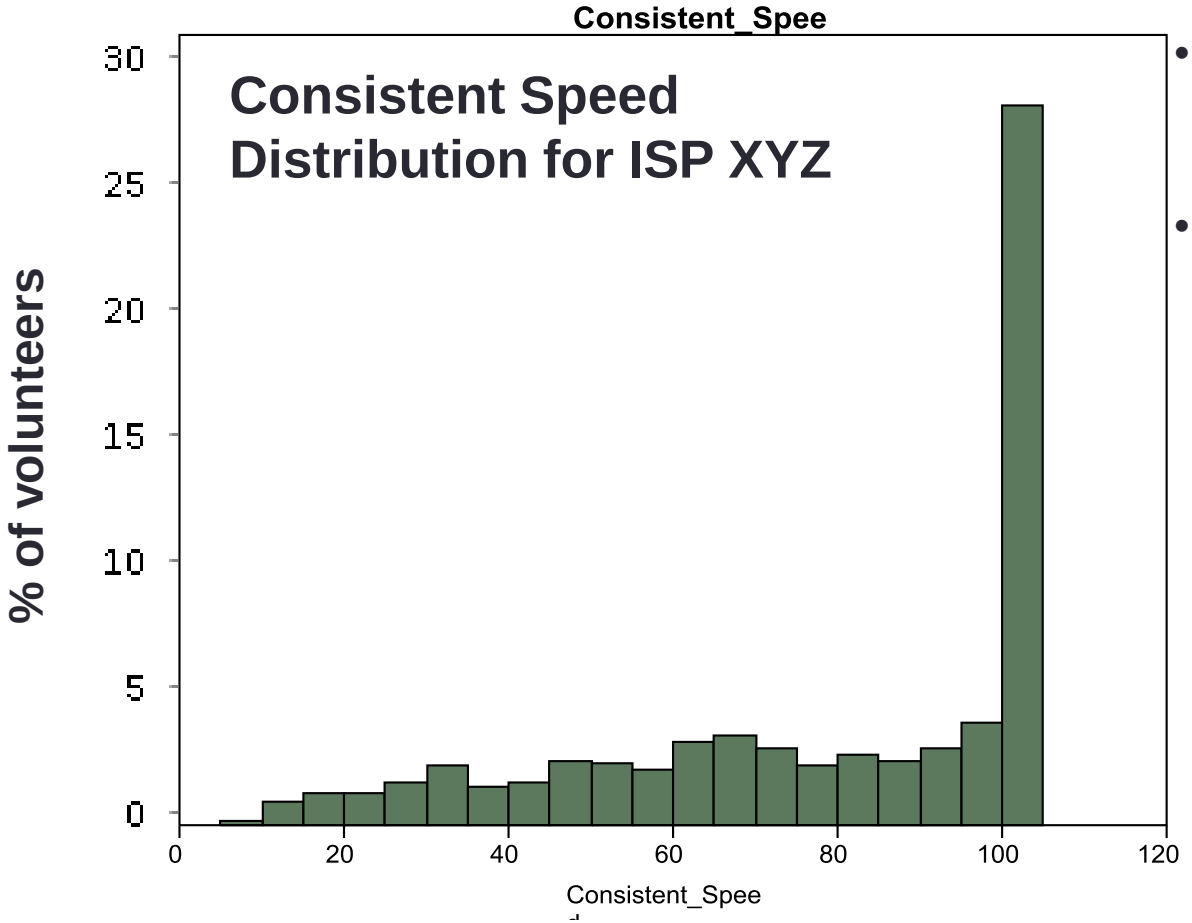


Statistical Methodology

- Challenges
  - The inferences drawn from the charts not supported by formal statistical tests
  - Outliers

- Suggestions:
  - Include the confidence intervals around the metric

Bootstrapping



- The consistent speed distribution is highly left skewed
- Needs bootstrapping to calculate confidence intervals

Building a Confidence Interval

99.99% Confidence Intervals for Median Consistent Speed				
ISP	State	Median	99.99% Confidence Interval	
ISP 1	CA	80	78	81
ISP 2	CA	83	83	83
ISP 3	CA	96	96	96
ISP 4	CA	99	99	100
ISP 5	CA	100	100	100
ISP 6	CA	91	76.5	100
ISP 7	CA	90.5	81	98.5
ISP 8	CA	98	98	98
ISP 9	CA	100	100	100
ISP 10	CA	100	97	100
ISP 11	CA	84	66	90

Conclusion

- Weekend prime time is the new Peak Period
- Consistent Speed – accounting for variance
- ISP alone does not tell the whole story
- Windows desktop application

Thank You

Q & A ?